Our first speaker this morning in this session is Dr. Homer Warner. I am sure all of us have known and worked with Dr. Warner. He and I both share membership on the Health Care Technology Study Section at this time. He is Professor and Head of the Department of Biophysics and Bioengineering at the University of Utah at Salt Lake City. He is Director of the Cardiovascular Activity and a Research Career awardee at NIH. Most of us know him for his creative and innovative work in the use of technology in Health Services.

Dr. Warner will discuss the topic, "Can Automation Make Interactive Medical History Taking Feasible and Acceptable?"

CAN AUTOMATION MAKE INTERACTIVE MEDICAL HISTORY TAKING FEASIBLE AND ACCEPTABLE?

Presented by Dr. Homer Warner
(Latter Day Saints Hospital)

DR. WARNER: I think the evaluation of the keynote speaker is probably best done by the number of questions he raises in the minds of his listeners. I would like to start off by disagreeing with Charlie on one of the first statements that he made. I hope that this will lead to a useful discussion.

I would like to tell you a story. We have been involved in a variety of computer applications in medicine over the last 10 years. One of these has been the automation interpretation electrocardiograms.

Alan Pryor in our department, who has done this work in ECG, went to Germany two years ago for the summer. While he was gone, he asked me if I would read the ECG's of all those patients going through the Admitting Screening Clinic and check them against the program.

During the month of July, I did that. I first looked at the electrocardiograms and made my interpretations and then I looked at what the computer said. The next month, I did it the other way around. I looked at what the computer said and then I looked at the electrocardiogram to see if I agreed with it. Until you have had that kind of experience, it is hard to appreciate the differences in the two procedures. One is an inductive process and the other is a deductive process. It is the difference between taking an essay test and taking a multiple choice test essentially.

When I said I disagreed with Charlie, what I am talking about is that I think the role of computers is not going to be primarily doing the drudgery, the data collection, but it is going to be altering the way we perform the intellectual tasks in medicine. Another experience that leads me to this bias, is related to the Intensive Care Monitoring that we have been engaged in for a number of years. In these Intensive Care wards,

we monitor the blood pressure and cardiac output and some other variables and present these to the nurse at regular intervals. We have alarm systems that alarm them about early trends in the patient's condition and so on. We thought we would do a great thing by relieving the nurse of the drudgery of counting the pulse and taking blood pressures. However, it turns out that nurses aren't happy--I am making a generality, it is not universally true --but more often than not, the nurse is more comfortable when she is around doing mechanical chores than when she is asked to do mental work. When the machine gives the alarm that there is a trend in the patient's condition, she now has to figure out what that means and make some decisions. She has a different kind of responsibility imposed upon her. She doesn't want to be alerted early in the game because it is too hard to interpret. The earlier in the process that you are aware that something may be going on, the more difficult it is to decide what the mechanism is. Downstream, after the patient is obviously in shock, there is no problem in deciding that he is probably bleeding internally. But early it is hard and the earlier, the harder. Not only that, but the earlier you detect something, the harder it is to bring it up out of the noise level. Therefore, we have shifted our emphasis across the whole spectrum of our activities toward not eliminating the mechanical chores, but tacking onto the mechanical chores which computers have to do a good number of intellectual chores, if you will, to try to change the process that the doctor or nurse is now asked to do from an inductive process to a deductive one. We have some evidence to date that this is being accepted and that it is helpful, although I won't imply that this is universally true either, or that we are getting universal acceptance at this stage of the game in our attempts to do this.

I would like to get off into one particular application in which this philosophy has been developed. That is in the area of history taking. As you know, the history is far and away the most important source of information about a patient. That sounds like a bold statement, but I don't think there are too many physicians who would disagree with that. We, at this point in time, are able to make 68 percent of the diagnoses that the physician makes on patients being admitted to the hospital just from the history alone. I put "make the diagnoses" in quotes because what I mean is we come up with a probability that we can attach to that particular diagnosis that raises it up out of the noise level and brings it to his attention. Not all these diagnoses are established in the sense that we confirm them downstream, but we can do that across the board in medicine right now.

There are not many other laboratory tests or sources of data that can even begin to compete with the history as a source of information about the patient. However, there are some problems with history as a source of information. First of all, in contrast to an automobile, for instance, which may fall apart quite suddenly and doesn't have a lot of built-in sensors to alert us as to when one of the parts is beginning to be modified structurally prior to the functional abnormality developing, the

body does have these sensors. These sensors, for the most part, are sensors that the patient can recognize as being something different from what he has experienced before. As a result, these are pretty effective alarm systems in most cases.

However, this again is not universally true. It is here that we begin to develop signal to noise problems, that is, that the individual is adaptive to those sensors.

As an example, it is not uncommon to find a patient who has been cyanotic all his life. He has had congenital heart disease and has been severely limited. Other people can look at him and recognize that he is having problems and yet, until he has been operated on and the defect has been repaired, he doesn't really know what normal is. He is amazed at how much better he feels.

I have a good friend that I have played tennis with who has an atrial defect. At about age 40, he had this repaired. He still could beat me before the operation, but he was amazed at the difference in his performance because he had become used to this symptom.

Therefore, in analyzing the data that a patient gives you from a history, we have to recognize that there are varying thresholds for almost any manifestation that the disease may have and that we are dealing with very noisy data. Because of that we have gone to and approached history taking quite differently than most other people working in this field who have used a branching binary tree-structured approach. We have used an approach based on conditional probability in which, at each decision node, the next question is chosen based on probabilistic knowledge rather than yes-no kind of prestructured decision logic.

I am going to assume that none of you know anything about this for purposes here this morning, because I recognize there probably is quite a diversity of background. Assume we have a table now in which we have diseases here and attributes here (Indicating). You call them symptoms. Under A, we will put an apriori probability for each of the diseases. This will be unique to each clinic. These probabilities should be a function of the disease itself and independent of the clinic.

If one starts with certain information about what the likelihood is of the patient having a particular disease when he walks into a clinic, he can then modify that probability after he receives each new piece of information. Using the Bayesian approach he can say a probability of the disease given the yes answer to the Jth question here is equal to the apriori probability of the disease and the probability of getting a yes answer to the Jth question. If he got a no answer, simply substitute one no to that probability. These are the numbers we have stored in here (Indicating). This has to be normalized by doing this across all diseases and summing it up. There are two assumptions that we are

making here. One is that the symptoms are independent of each other within a disease and the second is that the patient has only one of the diseases. I will show you how we get around that as we proceed.

The program operates as follows: the patient is first asked his age and sex and after the apriori probability has been modified by age and sex, the program then selects the question that has the most information in it. We define the most information by the product of two terms. First, we want the question which will minimize the entropy in the system. We say that the entropy is maximum when all diseases have equal probability. We want to find the question which will essentially raise one of the diseases out of the noise and bring it up to the maximum level. The second thing we want to ask is the question that has the highest probability of getting a yes answer. We certainly don't want to ask any questions that are not going to have a significant probability of getting a yes answer. We can find that by simply summing all the yes questions now. We look across all diseases and take the probability of that disease and times the probability of getting a yes answer to that question and the disease. This probability is the current probability that keeps getting modified, so that we now know the probability of getting a yes answer in the population as it now stands. Then we multiply that by this negative entropy term which is simply finding the question that will bring any one of the diseases to the highest probability.

Doing that, the computer then presents the next five best questions to the patient. The patient answers these questions either yes or no. Based on the answer, the probabilities are modified and we go through this until one of two things happens: either the information contained in the remaining questions, based on the probability as we have now modified them, falls below a critical level where we say there really are no questions worth asking that are going to help us further distinguish this, or we make a diagnosis. That is, we raise the probability of one of the diseases above 90 percent, at which time, we say there is no use pursuing this. We have enough to present it to the physician as a diagnostic suggestion.

At that time, we save that diagnosis. We now go back and set all the probabilities to their apriori level. After being modified by sex and age, we set them back to that initial level and they now go back and start a search again.

We did go through each system this way by setting the apriori for all diseases outside that system to zero just in the question selection phase, then they quickly switch back to their other values. This is the way you gain focus to get initial information on a system. Once you have one or more answers from a particular system, this will automatically lead you back into the full set of diseases.

You can make diagnoses even within the same system. That is part of the program.

The second part of that problem is to essentially generate this statistics matrix because, of course, all of the disease making is based on those numbers.

I really think we are at the phase now in medical computing where we have to talk about three elements: hardware, software and what I would like to call medicalware. By medicalware I simply mean things like this: the medical knowhow that makes that program meaningful, that to generate a matrix like this is in my opinion an iterative process. It is one where you have to first provide some initial approximation to get the program into the medical system in some way and to start collecting data. People are not going to let you collect data in an environment unless you are providing some feedback to the environment that makes it worthwhile.

What was done here was: I sat down with a cardiology man and a surgical man and between the three of us, we gathered as much information as we could to make a first approximation of a set of symptoms and diseases. We got something going. We then put that into the clinic and we screened 30 patients a day on the average, six days a week, at one of the hospitals.

We have three other hospitals on it now. I don't know what the load is on those hospitals. We have done this now for two years, so we have thousands of cases out there.

We quickly recognized,not only for this project but for others, that we had to have some kind of control to let you know downstream what the true answers were. Therefore, we had to introduce some kind of a cogent scheme to put the best final information into the system on each patient we could.

We developed a system based on systematized nomenclature of pathology. It was developed by the pathologist for coding pathological information. This uses four fields of information. There is a field to find the anatomical location of the disease process, what part of the body is involved. There is a field for the functional abnormality associated with it. There is a field for the nature of the disease process itself. There is a fourth field which defines the ideology or cause of the process. We have added a fifth field to that which is essentially a treatment field where we have all the surgical and operative codes and other kinds of treatments.

With that kind of a screen, you have not only a coding system, but a nomenclature that will allow you to enter words into the computer and have them coded automatically so that they are represented in the machine under one or more of these field codes and are retrievable in English text. This allows the doctor to describe the disease process on the front of the chart. A girl

can enter the terms in an interactive fashion. On an average of 30 seconds, she can enter that disease and have it coded. This allows us then to refine this matrix. We begin getting enough cases in there searching for a particular disease and asking ourselves, given that disease, how many of the patients, indeed, answered yes to this question and how many of those were asked the question. The record keeps a list to tell you which patients were asked a particular question and which patients answered yes to that question. We can begin collecting data and get the actual statistics from that. That has problems in itself, too.

A third mode we have had to generate this medical inquiry is to bring experts in and have them begin to look at the numbers we have in here and get their input. This has been a very informative thing. I think almost every area where we have tried to formalize the diagnostic process, we found that there are severe limitations in the reproducibility of what the physician does.

For instance, we took two experts and asked them to modify the statistics on a particular disease, how frequently does such and such a symptom appear in that disease? We will often get different opinions, so we then have to try to evaluate that in light of our current matrix. Therefore, it is an adequately adaptive process that we use. If we get an expert who gives us an opinion that differs greatly from what the existing opinion in there was, we think perhaps there is some compromise and we will go after a third opinion or arrive at a number that is a compromise between what we already had in there and what the new guess is.

With this we are gradually homing in on a set of numbers which represent fairly realistically the medical knowledge of today for these particular diseases. Also, we have developed some fairly sophisticated programs to allow you to look at this in a variety of ways. For instance, you can interact through the console to display back the probability of a particular disease listed by the most common symptoms first, and you can see these problems displayed, or you can look at a problem so that you list first that symptom which has the highest ratio of the probability in this disease versus this disease and you can see them. The first symptom we will list will be the one that is best for distinguishing these particular diseases. You can quickly spot inconsistencies here. It is important that you don't put zeros in very many of these spaces except in certain areas like sex. However, most of the other symptoms are not nearly that definitive.

Therefore, we are operating on fairly soft data and yet I feel that it is one of the more important areas that we must focus on because of the information content of that set of data.

I would like to display a slide as an example of how we print this out, because I think this is another important phase of the process of helping the physician.

This is a print-out of a history on a patient. First, it says that there is a history which suggests that the patient has an inguinal hernia because the patient is a male. This first lists the diseases in order of their probability. It says it is 100 percent likely because the patient is a male. He believes he has an inguinal hernia because he has a bulging, he is over 39 years of age, he has pain radiating to the groin and experiences a burning pain in his chest or abdomen. The most likely disease is printed first, then, in order of decreasing probability within a given disease, these symptoms are listed according to the most common need. One symptom is being a male with an inguinal hernia. This is the least helpful symptom and this is the most helpful one (Indicating).

We try to structure this in a way that makes it easier for the doctor to see what the data was that backed up that particular decision. It also suggests that the patient has hypertension because he has a relatively high blood pressure. He says he has high blood pressure. You might think that is a one to one thing, but it isn't. He feels weak and he gets tired easily. He has difficulty sleeping. He becomes flushed and has loss of strength on one side and so on. Then it is judged diabetes mellitus and angina pectorus and orthostatic hypotension, because the patient has difficulty breathing after a flight of stairs. But, that is not enough. This patient was admitted by a surgeon for a repair of an inguinal hernia. He got in. He did have hypertension. He has angina pectorus and orthstatic hypotension. He had an immediate family history of diabetes. He did not have it on follow-up. However, the surgeon, became sufficiently concerned over his age and his hypertension and he had an abnormal ECG so that he discharged him and did not repair his hernia.

Here is the rest of the history on this case (Indicating). It also goes through a systems review so that any of the questions answers that did not contribute to one of the diagnoses there appear under the systems review.

It took 13 minutes for the patient to complete that history. The average time is 11 minutes for a patient to go through that process.

When you talk about changing the intellectual task from an inductive one to a deductive one, this is what I am talking about. I am talking about essentially structuring the information based on some medicalware that is available in the computer in such a fashion that the physician's job is one of evaluation, essentially a multiple choice matter, rather than having to go from the total set of all probabilities and come up with the proper classification for that patient, which most of the subsequent decision making will depend on.

We have attempted to evaluate this history in its present form which, I think, is obvious from what I have said is not the ultimate form at all. It will improve as the numbers improve on which we are basing these decisions. We have tried to evaluate it not on the question of does it make appropriate decisions, but does it elicit proper answers to the questions.

First of all, we went through this by asking the physicians to fill in questionnaires about his patients. The physicians, however, did not report on each question. We had incomplete questionnaires. None reported that the patient had answered the questions incorrectly, that is, that the question answers given were inappropriate, although we know in fact that they were, in some cases, and although the physicians did not respond that way we had one patient in whom we diagnosed pancreatitis who was in to have a gall bladder removed. The question was asked of the patient if her gall bladder was removed. She misunderstood the question and answered yes, so of course we did not diagnosis gall bladder disease. There will be some of those, but that is not a major factor.

At the Latter Day Saints Hospital, which is a big general hospital, we have a better than average intellectual capacity. I think the level of literacy is higher than what you might find in some of the bigger cities in a large general hospital with mixed population. We are not focusing on that problem. That is a different kind of a problem than we have encountered.

DR. NICHOLS: Are these replies that you have here on a written form so that a person reads it and checks something, or is it verbal?

DR. WARNER: No, he reads it off the face of an oscilloscope.

DR. NICHOLS: And he presses a button for his answer?

DR. WARNER: Yes, each question, as it is presented, has an index by it, a number. Usually there are five at a time. There can be anywhere from one through eight questions at one time depending on the information content. However, in general, it is five, so these questions are numbered one, two, three, four, five. If he wants to answer two, three, and four, he hits two, three, four. When he wants the next set of questions, he hits another key and it brings up the next set of questions.

We have not had problems with the patient not understanding. We have a technician who initiates the process and watches the patient through the first set of questions to make sure he understands. Then, the technician goes off and leaves the patient. Sometimes a relative will have to come and help the patient, but we encourage the technician not to get involved any more than necessary in helping them interpret a question because then you have superimposed on that question the bias of the technician that happened to be there on that day.

The second thing we want to know is how accurate this is. I mentioned this number to you in reviewing the admitting diagnosis of the physician and comparing them against the list of diagnostic suggestions that are made by the computer. In 68 percent of the cases, the computer will have included in its differential the admitting diagnosis of the physician. That is not a great help to the physician, because he knows the admitting diagnosis. In general--not in the emergency situation--but where we help the physician is primarily in the secondary diagnosis that we will suggest. The case I showed you was a case in point. The surgeon was admitting the patient for inguinal hernia. The patient also had other medical problems that he was unaware of at the time of admitting. Subsequently, he became aware of those and, as a result, he changed his management of the patient.

We asked the following question to the physicians: Was the history helpful to you in this case? In 35 percent of the responses--these were questionnaires put with the history on the charts--the physician said yes, which is really a very high percentage if you stop and think about it because many of the cases were things that you cannot diagnosis from history. If a patient comes in to have a mole removed by a plastic surgeon, the things that we don't even try to diagnosis, it is like trying to diagnosis a dented fender on an automobile when the only symptom is a dented fender.

There is no logical process involved at all. That is true for a fair fraction of the medical problems that arise.

We asked the physicians: Has this been helpful to you on other cases? Twice as many answered yes to that question as "Was it helpful to you on this case?" This one could have predicted, I guess, because if we are going to help them we are not going to help them on every case. We are encouraged at this point that the physicians are accepting it and they are using it and this is the beginning point, this history of a problem-oriented approach to help with the management decision making of the patient. I will talk more about other aspects of this probabilistic and problem definition logic on Saturday.

MR. CAPLAN: I want to ask you a question about something you said very early in your talk with reference to your checking the automated ECG examination. You said at one point that you first looked at the ECG and then the computer printout, then you reversed the procedure. How did all this come out?

DR. WARNER: What do you mean? What is the present stage of the operation of that?

MR. CAPLAN: I believe you said you made a diagnosis, then you check the computer's result. Then you reversed the process. What was the result of that?

DR. WARNER: This was in a stage where we were doing the

comparison. This has been published since then. I can give you our numbers, but I want to emphasize that those numbers depend on the day's selection. If you are dealing with a population coming into the hospital, you will have the incidence of abnormal ECG's of approximately 20 percent. This is in our hospital. Our hospital is heavily oriented toward cardiovascular disease in the first place.

Again, we have done other series, evaluating the program on in-patients where we are limited to patients in whom the doctor for some reason expects that the patient has an abnormal ECG and, thus, orders one. Here you have a very high incidence--something like above 70 percent--that will be abnormal. If you look at the program's behavior in that population, it is not as good. It is a lot better at detecting normals from abnormals than it is in the specifics, where you are then asking the question: Did the computer agree with the consensus of two or more experts? That is what we use as our criterion in every respect on a diagnosis. We are at the present point with our program on that last category which is the most significant one, where you are dealing with the population of almost all abnormalities. We are at the level of 80 percent complete agreement between the physician and the consensus of the computer. This is essentially the same level as we get between agreement of two experts in the same ECG.

PROFESSOR WEED: You mentioned that in 35 percent of the cases, the medical people felt this was helpful. Could you explain this a little more? Was it helpful in that it reduced the time they put on it, or the cost, or their productivity, or was it just a confirmation?

DR. WARNER: All we did was essentially get some feeling that they thought it was a good thing on that case and in some way helped them with the management of the patient. We did not ask them what did you do differently as a result of this. The reason we did this was because I don't believe we would have gotten many responses. Doctors will not spend the time to give that kind of information. We have lots of anecdotal information on the cases that I have shown you here, but it is difficult to get.

DR. DWYER: I would like to ask Dr. Warner a question about the approach. He has the only system that has a probabilistic encounter model for taking the patient's history. The question is: Do you think--looking down the road and taking all the data you need for comparisons, if you want to--it is going to be possible to implement that system in a cost-effective manner?

DR. WARNER: We have implemented it in a cost-effective way. It costs us one dollar to take a patient's history. That does not mean you can set up a computer to do just history. However, for the fraction of the computer time used, the cost of the terminals, the personnel involved in doing it, we can do 30 histories a day from our Admitting Screening Clinic at the cost of a dollar apiece for the computer time used. It is cost-effective.

DR. DWYER: Do you feel that the probabilistic approach is a method that can be implemented or used in the sharpening of the art of differential diagnosis?

DR. WARNER: We hope it will do both. It is implemented and it is operating and it is self-supporting at this point. We have it going in four hospitals. All the experience I have described here is just from one. We have not fairly analyzed the others, but it is a useful service at this point. By the measures I have given you as accepted, beyond the stage of what does it do, what is it supposed to do, and is it accepted by the physician, the third is: How do you get a measure of what it is actually contributing to the practice of medicine? We don't have that data and it is difficult to get. It is awfully tough to get into that information process. Saturday, I am going to talk about how this fits in with the other parts of our decision-making-assist program that we are trying to implement in many other situations.

DR. SCHMITT: I think this fits into the discussion. One of the prime things we concluded at the Hanover Meeting on Automation of Electrocardiography was that the introduction of a continuum rather than a binary mathematical abnormality definition was important for the future automation of diagnosis. How does a continuum statement fit in with your probabilistic analysis? How would you hope to automate this?

DR. WARNER: We have published a prototype study on this. If you now consider a variable such as white blood count and perhaps the distribution of normal values would be from 5,000 to 10,000. This might be normally distributed for a normal population. For a patient with pneumonia, this may be skewed. For patients with leukemia, it is skewed differently. They are generally skewed toward the high side for that particular test. We found we could fit these distributions with a three-parameter curve. If you picture this matrix with disease and attributes here and white blood count in this cell, instead of storing one number which is the probability of the white blood count, you now store three numbers, which are the parameters that describe the distribution of that variable in the patient with that disease. Then you have a mean, a standard deviation and a skew factor. From this, we regenerate this curve and give particular white blood counts. Let's say we have a patient with a white blood count of 11,000. We can now go through and calculate the likelihood of getting a white blood count of 11,000 in a patient with that disease. That, then, plugs into the ba immediately you are right back into two other systems. We have published on that in the field of hematological disorders. This is a major undertaking. We have not done anything on it in the last year. Our clinical laboratory data has not been completed.

DR. SCHOEFFLER: One of the advantages of a chess playing program is that you can rate it numerically. Have you attempted to do anything like this here so that you can state diagnosis as well as a third-year medical student or an expert or any level

like this? That is, does it get all of the easy ones and none of the hard ones?

DR. WARNER: No, we have diagnosed some pretty sophisticated things like Addison's disease and hyperparathyroidism and myasthenia gravis. We tend to overdiagnosis these unusual things, which we think is not bad. It brings them to the attention of the physician. Most errors in diagnosis are not made by the possibilities under the set of circumstances. It includes many of the less common abnormalities. On the other hand, keep the set of diseases within limits. It is necessary to recognize that from the history alone, you cannot make fine distinctions between certain kinds of diseases. The symptoms are not that different. Therefore, we haven't tried to make the refinement in the diagnosis. That is going to depend upon laboratory tests that are sophisticated. We don't try to diagnose which coronary artery is cut, but we can, with a high degree of accuracy, diagnose the fact that you have coronary insufficiency. So the job is not as insurmountable as it might seem at first glance. The system is very flexible. However, we found that it stayed very stable. We have added some and we have taken some out. One of them we have in there right now is sarcoid. We find we are not diagnosing that. In fact, the logic will not let us diagnose it. You put in the symptoms of sarcoid and you come up with different things. I think the logical approach is to try to define diseases and say the history I suggest is tuberculosis or sarcoid.

PROFESSOR SEIREG: Has this been applied to infants?

DR. WARNER: No. This is all with adult populations. We have no pediatrics at our hospital.

DR. FLAGLE: Our second speaker this morning is Dr. Otto Schmitt, who is Professor of Biophysics and Bioengineering at the University of Minnesota. He is also a Professor of Physics, and the Director of a biophysical sciences group.

CAN AUTOMATION BROADEN THE SCOPE OF HEALTH SCREENING AND HEALTH MAINTENANCE?

Presented By Dr. Otto H. Schmitt
(University of Minnesota)

DR. SCHMITT: So that you will be warned about what I am trying to convey, there are two things. One of them is that the way in which we ask questions and the way in which we consider possible answers is tremendously influential on the answers we find. Therefore, I would like to put a good deal of attention on using technological methods in order to find diseases and go to the questions we can ask ourselves and the problems we pose, that is, the application of systems engineering thinking to the devising of problems to examine and consider as an unfamiliar and not very widely practiced task. As an amateur, I do it badly, but perhaps